

Complexity measures, emergence, and multiparticle correlations

Tobias Galla*

Theoretical Physics, School of Physics and Astronomy,
The University of Manchester, Manchester M13 9PL, United Kingdom

Otfried Gühne†

Naturwissenschaftlich-Technische Fakultät, Universität Siegen,
Walter-Flex-Straße 3, 57068 Siegen, Germany and
Institut für Quantenoptik und Quanteninformation,
Österreichische Akademie der Wissenschaften, Technikerstraße 21, 6020 Innsbruck, Austria

(Dated: April 3, 2012)

We study correlation measures for complex systems. First, we investigate some recently proposed measures based on information geometry. We show that these measures can increase under local transformations as well as under discarding particles, thereby questioning their interpretation as a quantifier for complexity or correlations. We then propose a refined definition of these measures, investigate its properties and discuss its numerical evaluation. As an example, we study coupled logistic maps and study the behavior of the different measures for that case. Finally, we investigate other local effects during the coarse graining of the complex system.

PACS numbers: 89.70.Cf, 89.75.Fb, 05.45.Ra, 05.65.+b

I. INTRODUCTION

The science of complex systems is an active area of research in physics and in adjacent disciplines. Many systems in physics, biology, the social sciences and in economics fall under its umbrella, and research into these topics is becoming increasingly interdisciplinary. The quantification of complexity, however, is not straightforward and there are many approaches to quantify complexity in more mathematical terms, see for example Refs. [1–8] or Ref. [9], where a number of complexity measures are listed and discussed.

A possible way to quantify the complexity of a correlated multiparticle system is to take its distribution in state space and consider its distance to the distribution of an uncorrelated system, measured for example by the multi-information or excess entropy [10, 11]. As recently demonstrated, this concept can be generalised to higher-order multi-particle correlations using methods of information geometry [10, 12, 13]. For a given N -particle system one can consider the space \mathcal{E}_k of N -partite distributions which are thermal states of k -particle Hamiltonians, and ask what the distance of a given N -particle distribution is from this space. These ideas were discussed in detail in Ref. [10], and an appropriate mathematical definition in terms of exponential families was devised. The authors of Ref. [10] have also made available an numerical algorithm with which to compute the complexity measures they introduce [14].

While the above mentioned approaches deal with *classical* complex systems, correlations play also a

vital role in *quantum* multi-particle systems. Consequently, many studies have been devoted to the characterization of quantum correlations [15–19]. Concepts from (classical) information geometry were extended to the quantum realm by Zhou [16, 18], in particular a quantum analog of the above exponential families \mathcal{E}_k was introduced. Based on these concepts the author of Ref. [16] arrives at the somewhat counterintuitive observation that local operations on a single particle can increase the overall correlation in a quantum state. More precisely, Zhou gives an explicit example of a thermal quantum state of a two-particle Hamiltonian on three particles, where a local operation on one of the three qubits converts the state into one which can no longer be written as a thermal state of a two-particle Hamiltonian, but which instead has three-particle correlations¹.

One of the main motivations for the present work is the question whether or not similar effects can be seen in classical complex systems. To this end we characterize the class of local transformations, acting on single particles, and investigate how multi-particle correlation, as defined in Ref. [10] is affected by those transformations. We show that (i) local transformations can turn a distribution generated by a two-particle Hamiltonian into one with genuine three par-

¹ In the following, statements such as ‘a state has k -particle interactions’ will always be taken to mean that the state cannot be written as the thermal (Gibbs) state of a Hamiltonian with $k-1$ -particle interaction alone, but that instead any Hamiltonian generating the state necessarily has at least one k -particle term. The absence of k -particle correlations in this sense is not equivalent to a factorization requirement on k -particle *correlation functions* often considered in statistical physics.

*Electronic address: tobias.galla@manchester.ac.uk

†Electronic address: ofried.guehne@uni-siegen.de

tile correlation, and as a corollary (ii) that integrating out individual particles can increase multi-particle correlation as well.

In order to provide a remedy for these undesired properties we propose the concept of a ‘local orbit’ of an exponential family. The local orbit of a set S of distributions is here the set of all distributions that can be generated from elements of S by applying local transformations. We investigate the possibility to define multi-particle complexity not in terms of distances from exponential families themselves, but from their local orbits, and we propose a numerical scheme with which these complexity measures can be approximated. Finally we discuss a second type of local intervention, and study how different coarse-graining procedures can affect the multi-particle correlation properties of a system of coupled chaotic maps.

The remainder of this paper is structured as follows: In Sec. II we provide the basic concepts of information geometry and briefly describe the complexity measures introduced in Ref. [10]. The focus of Sec. III is the concept of local transformations. We first introduce their mathematical definition, and then show how local transformations can increase the correlation content of multi-particle systems. In Sec. IV we propose an alternative set of complexity measures, invariant under local transformations. We also describe a numerical method with which to approximate these measures. In Sec. V we apply these concepts to a system of coupled chaotic maps, before we conclude in Sec. VI. Details of the numerical procedures are outlined in the Appendix.

II. COMPLEXITY MEASURES FROM INFORMATION GEOMETRY

In this section, we briefly review the complexity measures used in Ref. [10]. The underlying idea is the following: Consider a dynamical system composed of N particles each of which can be observed to be in one of two different states, referred to as 0 and 1 in the following. At any given time the state $\sigma = (\sigma_1, \dots, \sigma_N)$ of the system is therefore found to be an element of $\{0, 1\}^N$. The space of all 2^N possible configurations $\sigma = (\sigma_1, \dots, \sigma_N)$ will be denoted as $\Omega := \{(\sigma_1, \dots, \sigma_N), \sigma_i \in \{0, 1\}, i = 1, \dots, N\} = \{0, 1\}^N$. The stationary distribution of the system is then a probability distribution over Ω .

Given such a distribution, $P(\cdot)$, one can ask whether or not it is the thermal state of k -particle Hamiltonian, $0 \leq k \leq N$, i.e., whether one can write $P(\sigma) = Z^{-1} \exp[H^{(k)}(\sigma)]$, where $H^{(k)}(\sigma)$ is a Hamiltonian containing only j -particle terms with $j = 1, \dots, k$, and where Z is a constant ensuring normalization. Trivially, any N -particle distribution can be written as the Gibbs state of an N -particle Hamiltonian, but for $k < N$ this may not be the case. One

can then use the distance D_k (quantified by the relative entropy) from the set of all distributions generated by k -particle interactions as a measure of complexity. In order to make this idea mathematically precise, we first give the definition of exponential families, their closure, and the relative entropy. We here follow the work and notation of Ref. [10]. We then discuss some properties of the resulting distance measures in Sec. II B.

A. Exponential families

We will write the set of all particles as $V = \{1, \dots, N\}$ in the following. Given a subset $A \subset V$, we write $H_A(\cdot)$ for Hamiltonians (or, more generally, arbitrary functions) which only depend on the $\{\sigma_i\}$ with $i \in A$, i.e. $H_A = H_A(\{\sigma_i\}_{i \in A})$. A k -particle Hamiltonian ($k \leq N$) is then a Hamiltonian containing only terms which each depend on (at most) k spins, so it is of the form

$$H(\sigma) = \sum_{A \subset V: |A|=k} H_A(\{\sigma_i\}_{i \in A}). \quad (1)$$

We note that any k -particle Hamiltonian is also an ℓ -particle Hamiltonian for all $\ell \geq k$. The vector space of all k -particle Hamiltonians will be denoted by \mathcal{Q}_k in the following, and accordingly we have $\mathcal{Q}_k \subset \mathcal{Q}_\ell$ for all $\ell > k$.

The set of thermal states (or Gibbs measures) generated by k -particle Hamiltonians then constitutes a so-called exponential family. We write

$$\mathcal{E}_k := \left\{ P \mid P(\sigma) = \frac{e^{H(\sigma)}}{\sum_{\sigma'} e^{H(\sigma')}}, H \in \mathcal{Q}_k \right\} \subset \mathcal{P}(\Omega), \quad (2)$$

where $\mathcal{P}(\Omega)$ denotes the set of all probability distributions over Ω . We then have

$$\mathcal{E}_1 \subsetneq \mathcal{E}_2 \subsetneq \dots \subsetneq \mathcal{E}_N = \mathcal{P}(\Omega). \quad (3)$$

We will first discuss some of the properties of the families \mathcal{E}_k . First, it is important to note that the probability distribution in \mathcal{E}_1 are simply product distributions factorizing over single particles, i.e. they are distributions which can be written as $P(\sigma) = P^{(1)}(\sigma_1)P^{(2)}(\sigma_2) \cdot \dots \cdot P^{(N)}(\sigma_N)$ with single-particle distributions $P^{(i)}(\cdot)$. The distributions in the more complex exponential family \mathcal{E}_2 on the contrary can be written as products of the form

$$P(\sigma) = \mathcal{N} \prod_{i,j \in V, i \neq j} P^{(ij)}(\sigma_i, \sigma_j), \quad (4)$$

where the $\{P^{(ij)}\}$ are two-particle distributions and where the constant \mathcal{N} provides appropriate normalization. Factorizations of this type can be extended straightforwardly to describe the elements of \mathcal{E}_k as

products of k -particle distributions. There are other equivalent characterizations of \mathcal{E}_k , we will discuss and use some of them later in Section III B.

Given that Hamiltonians assign finite energies to all configurations σ , i.e. $H(\sigma) \in \mathbb{R}$ for all $H \in \mathcal{Q}_k$, $k = 0, \dots, N$ and all $\sigma \in \Omega$ the probability distributions \mathcal{E}_k by construction carry non-vanishing probability for all $\sigma \in \Omega$, i.e. $P(\sigma) > 0$ for all σ and all $P \in \mathcal{E}_k$. It is therefore useful to consider the closure $\overline{\mathcal{E}}_k$ to incorporate probability distributions without full support. For example, the three-particle probability distribution P with $P(000) = P(111) = 1/2$ and $P(\cdot) = 0$ elsewhere is not in \mathcal{E}_2 , but in $\overline{\mathcal{E}}_2$, since it can be approximated by the low-temperature limit of an Ising-type two-particle interaction. The generalization of this distribution to more than three particles can still be generated with two-particle interactions only. On the other hand, the N -particle distribution P with

$$P(\sigma) = \begin{cases} 1/2^{N-1} & \text{if } \sum_i \sigma_i \equiv 0 \pmod{2}, \\ 0 & \text{elsewhere} \end{cases} \quad (5)$$

cannot be generated via $N-1$ -particle interactions, but instead requires a Hamiltonian with N -particle interaction. This distribution assigns probability $2^{-(N-1)}$ to any bitstring σ with an *even* number of bits equal to one, and zero to all other bitstrings. If only $N-1$ entries in σ are known, it is not possible to tell whether or not this condition is met, hence one cannot decide whether $P(\sigma) = 2^{-(N-1)}$ or whether $P(\sigma) = 0$, this requires always knowledge of the state of all particles [see also point (iii) below].

B. Distance measures

Given a probability distribution $P(\cdot)$ over Ω and an interaction order $k = 0, \dots, N-1$, we then ask how closely $P(\cdot)$ can be approximated by distributions generated by k -particle Hamiltonians. This is captured by the following ‘distance’

$$D_k(P) := \inf_{Q \in \overline{\mathcal{E}}_k} D(P||Q), \quad (6)$$

where $D(P||Q)$ is the Kullback-Leibler distance [20]

$$D(P||Q) = \sum_{\sigma \in \Omega} P(\sigma) \log_2 \frac{P(\sigma)}{Q(\sigma)}. \quad (7)$$

This quantity has the following properties:

- (i) Since $D(P||Q) = 0$ if and only if $P = Q$, the quantity $D_k(P)$ is non-vanishing for all distributions $P \notin \overline{\mathcal{E}}_k$.
- (ii) The minimizing distribution $Q^* \in \overline{\mathcal{E}}_k$ on the RHS of Eq. (6) is given by the maximum likelihood approximation of P by distributions in $\overline{\mathcal{E}}_k$, see Refs. [10, 12] for further details.

(iii) A second characterization of the minimizer Q^* is the following: Q^* is the distribution of maximal entropy in the set of distributions with the same k -particle marginals as P [12]. This implies that the distribution from Eq. (5) has $D_k(P) = 1$ for all $k < N$, since the flat distribution has the same $k-1$ -particle marginals as P and clearly maximizes the entropy over all distributions.

(iv) For the case $k = 1$ the minimizer Q^* can directly be found as

$$Q^*(\sigma) = P^{(1)}(\sigma_1)P^{(2)}(\sigma_2) \cdot \dots \cdot P^{(N)}(\sigma_N), \quad (8)$$

where the $P^{(i)}(\cdot)$ are the single-particle marginals of P . The quantity D_1 is also known as multi-information [12, 21], the above quantities D_k can be considered as a generalization of this concept.

(v) For $k \geq 2$ an analytical calculation of $D_k(P)$ for a given distribution $P \in \mathcal{P}(\Omega)$ is usually not straightforward. There are, however, powerful numerical tools for its computation [14], for completeness we outline a possible algorithm in the Appendix.

(vi) Following Ref. [10] one can also consider

$$I^{(k)}(P) = D_{k-1}(P) - D_k(P), k = 1, \dots, N \quad (9)$$

as complexity measures. The quantity $D_k(P)$ here represents the *improvement* in approximating $P \in \mathcal{P}(\Omega)$ when k -particle terms in the generating Hamiltonian are allowed over the case in which only $k-1$ -particle interaction is admitted. In this paper, however, we will mainly work with $D_k(P)$.

(vii) Finally, possible generalizations to the quantum setting have been discussed in Refs. [16–18].

We also note that the choice of the Kullback-Leibler divergence [Eq. (6)] is of course not the only choice of an underlying distance between probability distributions. Other distance measures are conceivable, see for example Refs. [6–8]. In our work we will however restrict the discussion to the relative entropy, as this provides several useful properties of $D(\cdot||\cdot)$ which we will use later on.

III. LOCAL TRANSFORMATIONS

A. Definition and interpretation

We will next discuss the behavior of the measure D_k under local transformations. In considering local transformations we have the following scenario in mind (see also Fig. 1): a complex system composed of

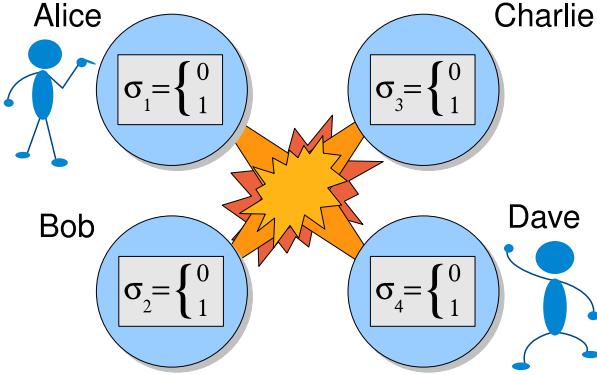


FIG. 1: (color online). Illustration of the effect of local transformations: A complex interacting dynamics generates a probability distribution over the state space of four binary particles. Each observer (Alice, Bob, Charlie and Dave) has access to the state of one particle, and reports these to a central authority (not shown). One type of local transformation arises, if one party, say Alice, changes the information she transmits stochastically, but without communication with the other observers (see text for further details). A second possible type of local transformation is given by disregarding one party, say Dave, and considering only the marginal distribution on the remaining three parties (this corresponds to a case in which Dave randomly reports to have seen state 0 or 1 regardless of what he actually observed). It is natural to require from a measure of complexity that it should not increase under either type of local transformations.

N binary particles generates successive states $\sigma(t) = [\sigma_1(t), \sigma_2(t), \dots, \sigma_N(t)]$, where $t = 1, 2, \dots$ denotes time. The resulting stationary distribution² over $\Omega = \{0, 1\}^N$ is then given by the cumulative histogram of all observed N -particles states in the long-time limit, i.e. $P(\sigma) = \lim_{t \rightarrow \infty} t^{-1} \sum_{t'=1}^t \delta_{\sigma\sigma(t')}$, where δ is the standard Kronecker symbol, i.e. $\delta_{\sigma\sigma'} = 1$ if $\sigma = \sigma'$ and $\delta_{\sigma\sigma'} = 0$ otherwise. Computing this distribution however requires a centralized observer, who has access to the sequence $\sigma(t)$.

To introduce the concept of local transformations we now consider a different scenario and assume that there are N distant observers who each have access to sequence of states of only one of the N particles, and then report these to a centralized agent who aggregates the information. By local transformations we mean manipulations a single observer can execute on the state of the particle he (or she) has observed before passing on the information to the centralized site. Assume for example that observer $i \in \{1, \dots, N\}$ upon observation of state 0 reports to have seen state 0 with probability $1 - a_i$ and that he (or she) reports

state 1 with probability a_i in such circumstances. Similarly if presented with observation 1 he or she may report state 1 to the central site with probability $1 - b_i$, but pass on wrong information with rate b_i ($a_i, b_i \in [0, 1]$). The central site then accumulates the information and obtains an N -particle distribution $\tilde{P}(\cdot)$, computed from the reported data of the N observers.

Unless P itself is pathological one will generally have $P = \tilde{P}$ only if $a_i = b_i = 0$ for all i . Behaviour on the part of the observers in which this is not the case can be thought of as noise (i.e. inaccurate observation) or as deliberate falsification of the observed input. The $\{a_i, b_i\}$ are error rates, quantifying how frequently inaccurate information is transmitted to the central agent. The crucial point is that such manipulations occur locally, i.e. different observers may not collude with each other and communicate before reporting to the central site. All operations therefore act on single particles only. It appears reasonable to ask that such transformations should not increase the complexity of $P(\cdot)$: if under a given measure of complexity k -particle correlations are absent in $P(\cdot)$ then applying local transformations should not introduce k -particle correlation in \tilde{P} (the effective distribution perceived by the aggregating central agent). In a similar vein disregarding one or several particles should not introduce additional correlation: if one considers the marginal probability distribution of the states of a subset of $M < N$ particles, one may require that these M -particle marginals should be not more complex than the joint probability distribution of all N particles. Similar considerations are well known from quantum information theory, where a basic requirement for entanglement measures is that they should not increase under local operations and classical communication.

In order to investigate whether the complexity measures defined above do indeed respect these constraints we write $A_{\mu_i, \nu_i}^{(i)}$ for the probability with which observer i reports state $\mu_i \in \{0, 1\}$ to the central site when he/she has in fact seen state $\nu_i \in \{0, 1\}$. We always have $\sum_{\mu_i} A_{\mu_i, \nu_i}^{(i)} = 1$ for all i and ν_i . We will use the notation $A^{(i)} = \begin{pmatrix} 1 - a_i & b_i \\ a_i & 1 - b_i \end{pmatrix}$ in the following. A specific local transformation is then characterized by the parameters $\omega = (a_1, b_1, \dots, a_N, b_N) \in [0, 1]^{2N}$, we will often use T_ω as a short-hand for the transformation defined by ω .

For any local transformation $T^{\text{loc}} = T_\omega$ the resulting distribution $\tilde{P}(\cdot)$ put together by the central site can then be written as

$$\tilde{P}(\mu) = \sum_{\nu} T_{\mu\nu}^{\text{loc}} P(\nu), \quad (10)$$

where

$$T^{\text{loc}} = \bigotimes_{i=1}^N A^{(i)} = \bigotimes_{i=1}^N \begin{pmatrix} 1 - a_i & b_i \\ a_i & 1 - b_i \end{pmatrix}. \quad (11)$$

² We here only consider interacting systems which do have a stationary distribution.

Equivalently the probability of the central site constructing the N -particle configuration μ from the reports he or she receives from the N observers when the true configuration of the N -particle system was ν is given by

$$T_{\mu\nu}^{\text{loc}} = \prod_{i=1}^N A_{\mu_i\nu_i}^{(i)}. \quad (12)$$

The superscript ‘loc’ here indicates that we are interested in local transformations, i.e. those for which the matrix T has the product structure indicated in Eqs. (11, 12). More generally we can consider linear transformations $Q(\mu) = \sum_{\nu} T_{\mu\nu} P(\nu)$. These are well defined so long as the entries of the matrix $T_{\mu\nu}$ are nonnegative and the entries in the columns add up to one, $\sum_{\mu} T_{\mu\nu} = 1$, i.e. so long as the matrix T is stochastic³.

In order to probe the measures of correlations introduced earlier we next investigate how the distances D_k behave under local operations. As it turns out, distances from exponential families can indeed increase under local transformations.

B. Local operations can increase D_k

To give an example, first note that under arbitrary transformations (not necessarily local) the Kullback Leibler distance can only decrease, i.e., $D(P||Q) \geq D[T(P)||T(Q)]$ [20], where we have written $T(P)$ for the image of $P(\cdot)$ under a transformation mediated by the matrix T [and similar for $T(Q)$]. As a consequence, the following two statements are equivalent:

- (1) D_k decreases under local transformations, i.e. $D_k[T^{\text{loc}}(P)] \leq D_k(P)$ for all P and all local transformations T^{loc} ,
- (2) the manifold \mathcal{E}_k is invariant under local transformations, i.e. $T^{\text{loc}}(Q) \in \mathcal{E}_k$ for all $Q \in \mathcal{E}_k$ and all T^{loc} .

As we will now show neither of these two statements is true however.

In order to construct a counterexample it is useful to detail further the characteristics of probability distributions in \mathcal{E}_k . For simplicity, we focus on the case of three particles and consider \mathcal{E}_2 , generalization to the other cases is straightforward. It turns out to be helpful to formulate the problem in the language

of quantum mechanics. In analogy to the space consisting of the 2^3 configurations $\{\sigma\} = \{(\sigma_1, \dots, \sigma_3)\} = \{(0, 0, 0), (0, 0, 1), \dots, (1, 1, 1)\}$ we consider analogous quantum states in the orthonormal computational basis, $\{|\sigma\rangle\} = \{|000\rangle, |001\rangle, \dots, |111\rangle\}$. For a given probability distribution $P(\sigma)$ we then consider the Hamiltonian defined by

$$H = \sum_{\sigma} \lambda_{\sigma} |\sigma\rangle\langle\sigma|, \quad (13)$$

where the $\{\lambda_{\sigma}\}$ are given by $\lambda_{\sigma} = \ln P(\sigma)$. For the moment we restrict the discussion to probability distributions P with full support, i.e. we assume $P(\sigma) > 0$ for all σ . The operator H is then well defined and diagonal in the computational basis, $H|\sigma\rangle = \lambda_{\sigma}|\sigma\rangle$, and we have

$$P(\sigma) = \langle\sigma| \exp(H)|\sigma\rangle. \quad (14)$$

Using the fact that tensor products of Pauli matrices form a basis of the space hermitian operators, any Hamiltonian of the type defined in Eq. (13) can uniquely be written in the form

$$H = \kappa \mathbb{1} + \sum_i \alpha_i \sigma_z^{(i)} + \sum_{i,j} \beta_{ij} \sigma_z^{(i)} \otimes \sigma_z^{(j)} + \gamma \sigma_z^{(1)} \otimes \sigma_z^{(2)} \otimes \sigma_z^{(3)}, \quad (15)$$

where $\sigma_z^{(i)}$ denotes the Pauli matrix σ_z acting on qubit i . We will in the following generally not treat κ in Eq. (15) as free parameter, but choose it such that the distribution $P(\cdot)$ constructed from H via Eq. (14) is correctly normalised.

A Hamiltonian of the type described in Eq. (15) is a two-particle Hamiltonian if and only if the 3-particle term is absent, i.e. if $\gamma = 0$. This is the case if

$$\text{tr} \left[H \left(\sigma_z^{(1)} \otimes \sigma_z^{(2)} \otimes \sigma_z^{(3)} \right) \right] = 0. \quad (16)$$

This in turn is equivalent to $\lambda_{000} + \lambda_{011} + \lambda_{110} + \lambda_{101} = \lambda_{001} + \lambda_{010} + \lambda_{100} + \lambda_{111}$. So any probability distribution $P \in \mathcal{E}_2$ fulfils $\ln[P(000)] + \ln[P(011)] + \ln[P(110)] + \ln[P(101)] = \ln[P(001)] + \ln[P(010)] + \ln[P(100)] + \ln[P(111)]$, see also Ref. [12]. Equivalently, one has

$$\begin{aligned} & P(000) \cdot P(011) \cdot P(110) \cdot P(101) \\ & = P(001) \cdot P(010) \cdot P(100) \cdot P(111). \end{aligned} \quad (17)$$

These considerations can now be used to show that the above two statements (1) and (2) do not hold, and that instead local transformations do not leave \mathcal{E}_2 invariant. To this end one use Eq. (15) to generate two-particle Hamiltonians and their associated probability distributions $P \in \mathcal{E}_2$ at random and then subsequently apply a local transformation T^{loc} with randomly chosen matrix elements $A_{\mu_i, \nu_i}^{(i)}$. Relation

³ If one requires in addition that the row sums are normalized, $\sum_{\nu} T_{\mu\nu} = 1$ the matrix is called doubly stochastic and the entropy increases during the process. Doubly stochastic matrices naturally occur, if one requires that the maximally mixed distribution is invariant under the transformation.

(17) can then be used to check whether or not the outcome is still in \mathcal{E}_2 . It turns out that generally $T^{\text{loc}}(P) \notin \mathcal{E}_2$. This proves that D_2 can increase under local operations.

Our considerations complement those of Ref. [16], where it was observed that a quantum analog of the quantity D_k can increase under local operations and classical communication. In the quantum case, however, the possible local transformations form a much larger class of maps: For instance, the positive maps on a single qubit are parameterized by 12 parameters, while in the classical case, a transformation on a single particle has only two degrees of freedom. Our example shows that the increase of D_k occurs already in the classical regime and is also not due to possible non-commuting terms in the Hamiltonian in the quantum setting.

Finally, we would like to stress that the manifold \mathcal{E}_1 of distributions factorizing over individual particles is clearly invariant under local transformations. Hence, the quantity D_1 (also referred to as multi-information [12, 21]) does not increase under local transformations.

C. Tracing out particles can increase D_k

A second requirement one may ask of correlation measures is that they do not increase when individual degrees of freedom are integrated out. Mathematically formulated, the question is whether the M -particle marginals ($M < N$) of a N -particle distribution in \mathcal{E}_k are still generated by a k -particle Hamiltonian. For our later discussion it is important to realize that tracing out individual particles can be understood as carrying out a local transformation with specific parameters. More precisely, one may for example choose to apply the local transformation defined by

$$A^{(1)} = \begin{pmatrix} 1 - a_1 & a_1 \\ a_1 & 1 - a_1 \end{pmatrix} \quad (18)$$

and

$$A^{(k)} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad (19)$$

for $k = 2, \dots, N$. For the first particle reports state 0 to the central site with probability $1 - a_1$, and state 1 with probability a_1 irrespective of the actual state of the particle. The states of the remaining particles are reported faithfully. If applied to an N -particle distribution $P(\cdot)$ the transformation $T^{\text{loc}} = \bigotimes_{i=1}^N A^{(i)}$ will result in a distribution of the form

$$T^{\text{loc}}(P)(\sigma) = R(\sigma_1)P'(\sigma_2, \dots, \sigma_N) \quad (20)$$

where $R(\sigma_1 = 0) = 1 - a_1$ and $R(\sigma_1 = 1) = a_1$, and where $P'(\cdot)$ is the $N - 1$ -particle marginal of

the original distribution $P(\cdot)$, i.e., $P'(\sigma_2, \dots, \sigma_N) = \sum_{\sigma_1} P(\sigma_1, \sigma_2, \dots, \sigma_N)$. Thus $T^{\text{loc}}(P) \in \mathcal{E}_k$ if and only if $P' \in \mathcal{E}_k$.

Again, one can directly find counterexamples, which show that D_k can increase under tracing our particles. For instance, if one generates a random four-particle distribution in \mathcal{E}_2 , then, after tracing out one particle, the marginal distribution is generally not found to be in \mathcal{E}_2 .

IV. INCORPORATING LOCAL TRANSFORMATIONS IN THE CORRELATION MEASURE

In the last section we have seen that the measures D_k can increase under local transformations, and that they are hence lacking a desirable property of correlation measures. To overcome this, we may replace the manifolds \mathcal{E}_k by their local orbits,

$$\mathcal{L}_k = \{T(P) | T \in \mathcal{T}^{\text{loc}}, P \in \mathcal{E}_k\}. \quad (21)$$

The notation \mathcal{T}^{loc} here indicates local transformations of the type defined in Eqs. (11,12). The set \mathcal{L}_k is manifestly invariant under local transformations, hence the quantity

$$C_k(P) = \inf_{Q \in \mathcal{L}_k} D(P \| Q) \quad (22)$$

is a correlation measure which is nonincreasing under local transformations. It is also nonincreasing under tracing out particles, as this is a special case of a local transformation. Given the invariance of \mathcal{E}_1 under local transformations, the quantity C_1 coincides with the multi-information.

In the following we will first discuss some basic properties of this measure, before we will describe how to approximate C_k numerically. In the next session we will then apply this measure of complexity to the example of a set of coupled chaotic maps, we will also provide a comparison with existing results on the measure D_k [10].

A. Properties of the set \mathcal{L}_k and the measure C_k

Let us first discuss some properties of \mathcal{L}_k . First we observe that for $N > 4$ particles \mathcal{L}_2 is a set of measure zero in the space of all probability distributions. To see this one first notes that \mathcal{E}_k is characterized by $\sum_{i=1}^k \binom{N}{i}$ free parameters, this is the number of possible terms in a Hamiltonian up to and including order k (see Section III B). Furthermore, local transformations are parameterized by two variables per particle, i.e. by $2N$ parameters in total. Given that N -particle distributions carry $2^N - 1$ free parameters and that $2^N - 1 > \sum_{i=1}^2 \binom{N}{i} + 2N$ for $N \geq 5$ we conclude that

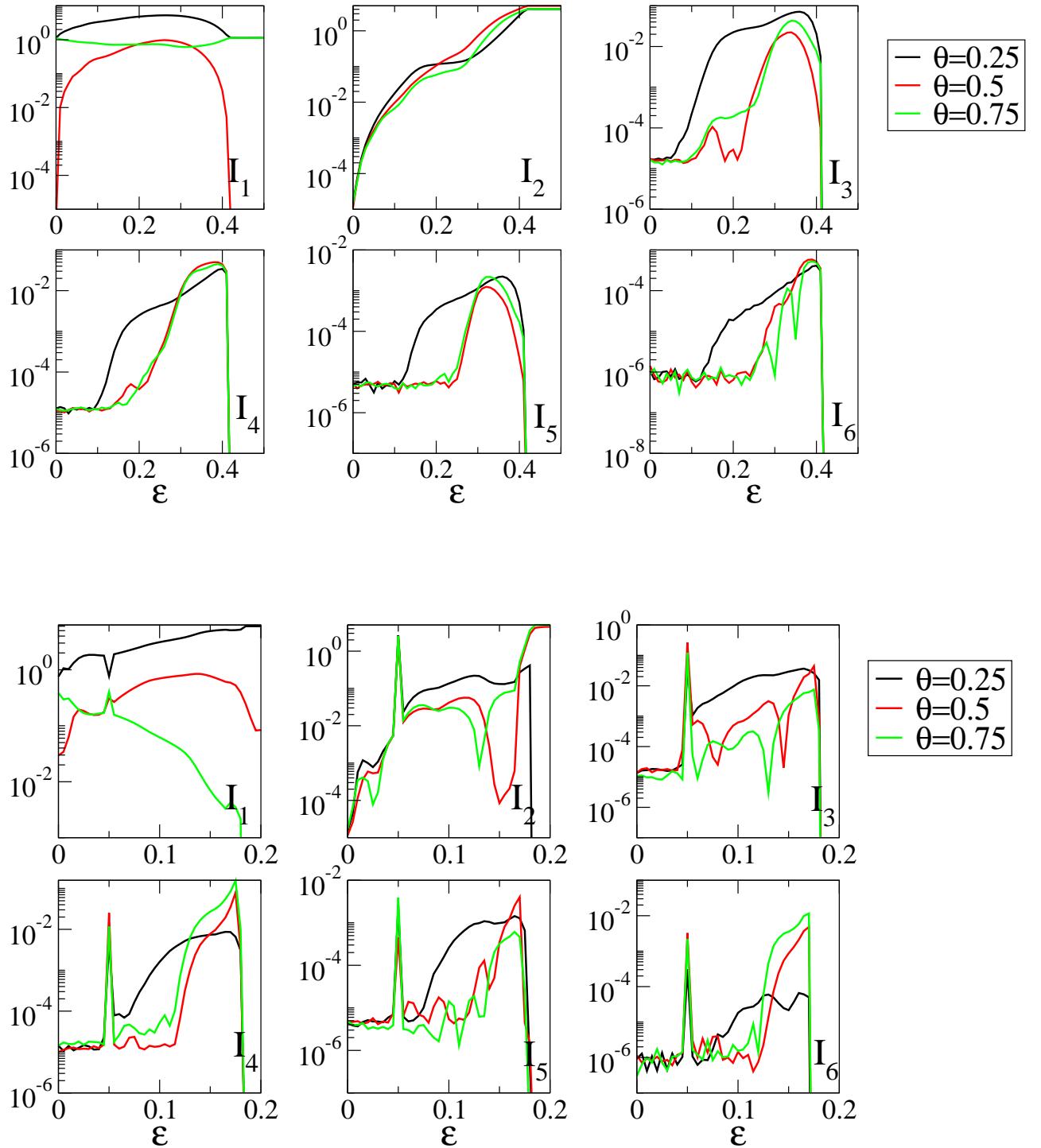


FIG. 2: (color online). Complexity measures I_k , $k = 1, \dots, 6$ as a function of the coupling parameter ε for a system of $N = 6$ coupled tent maps (upper six panels), and for $N = 6$ coupled logistic maps (lower six panels). Each panel shows three curves corresponding to coarse-graining procedures with thresholds $\Theta = 0.25, 0.5$ and 0.75 respectively. All data is obtained from running 1000 iterations of the iterative projection algorithm. Curves represent the average of I_k over 20 samples with independent random initial conditions $x_i(t = 0) \in [0, 1]$.

the set \mathcal{L}_2 is of measure zero in the space of all probability distributions. Similarly, one can argue that the \mathcal{L}_k for small k are of measure zero for many particles.

For lower particle numbers the set \mathcal{L}_2 does not

cover the whole space of distributions either. To see this, consider three particles and the probability distribution $P(\cdot)$ with $P(000) = P(011) = P(101) = P(110) = 1/4$, and $P(\sigma) = 0$ otherwise. We know al-

ready that $D_2(P) = 1$ is relatively large, hence it is a natural candidate to be outside of \mathcal{L}_2 . In order to show that this is indeed the case, assume there exists a local transformation $T \in \mathcal{T}^{(\text{loc})}$ and a $Q \in \mathcal{E}_2$ such that $P = T(Q)$. We parameterize $T = A^{(1)} \otimes A^{(2)} \otimes A^{(3)}$ via $A^{(k)} = \begin{pmatrix} 1-a_k & b_k \\ a_k & 1-b_k \end{pmatrix}$ and we have

$$\begin{aligned} P(001) &\geq Q(000)(1-a_1)(1-a_2)a_3, \\ P(010) &\geq Q(000)(1-a_1)a_2(1-a_3), \\ P(100) &\geq Q(000)a_1(1-a_2)(1-a_3), \\ P(111) &\geq Q(000)a_1a_2a_3. \end{aligned} \quad (23)$$

The distribution $P(\cdot)$ is constructed such that the left-hand-sides of these equations vanish. Assume now that $Q(000) > 0$. In order to satisfy the above inequalities we must have $a_k \in \{0, 1\}$ for all k . In a similar manner, making the assumption $Q(111) > 0$ leads to the requirement $b_k \in \{0, 1\}$ for all k .

For cases in which $Q(000)$ and $Q(111)$ are both non-zero this solves the problem, this includes in particular all distributions $Q(\cdot)$ with full support. The local transformations are then such that any party applies one of the following transformations: (i) they do not make any modification to the state they observe ($a_k = b_k = 0$), (ii) they always flip the binary symbol they receive ($a_k = b_k = 1$) or (iii) they always report the same result to the central site no matter what input they receive (e.g., if $a_k = 0, b_k = 1$ they always report 0 and if $a_k = 1, b_k = 0$ they always report 1). With these transformations, however, one cannot obtain the probability distribution $P(\cdot)$ from a distribution in \mathcal{E}_2 [in the case (iii) this is impossible as P assigns positive probability to bit strings in which the variable σ_k takes either of the two values 0 and 1].

It remains to address cases for which there are $\sigma \in \Omega$ such that $Q(\sigma) = 0$ (i.e. Q is in the closure $\overline{\mathcal{E}_2}$, but not in \mathcal{E}_2 itself). First, using the reasoning from above, one can directly see that if $Q(0, \sigma_2, \sigma_3) > 0$ for some σ_2, σ_3 , then $a_1 \in \{0, 1\}$ and if $Q(1, \sigma_2, \sigma_3) > 0$ for some σ_2, σ_3 , then $b_1 \in \{0, 1\}$. This solves the problem for many further cases, for instance if there are two events $\sigma_1, \sigma_2, \sigma_3$ and $\sigma'_1, \sigma'_2, \sigma'_3$ with $\sigma_k \neq \sigma'_k$ and $Q(\sigma_1, \sigma_2, \sigma_3) > 0$ as well as $Q(\sigma'_1, \sigma'_2, \sigma'_3) > 0$.

The only remaining cases, where we cannot set any constraints on a_1 (or b_1) are of the type where $Q(0, \sigma_2, \sigma_3) = 0$ for all σ_2, σ_3 [or $Q(1, \sigma_2, \sigma_3) = 0$ for all σ_2, σ_3]. But for such cases the value of a_1 (or b_1) has no effect on $T(Q)$, so, without changing the outcome of the transformation, such a parameter can be set to either 0 or 1, again leading us to the desired result. Finally, note that due to continuity reasons and the fact that $\overline{\mathcal{E}_2}$ is compact, P has a finite distance to \mathcal{L}_2 . Using the algorithm outlined in the next section we find a numerical distance of $C_2(P) \approx 0.689$.

B. Numerical calculation of C_k

It is not a-priori not straightforward to compute $C_k(P)$ for a given N -particle distribution $P(\cdot)$. We have however made some progress in devising an iterative numerical scheme. We cannot prove at the moment that it always converges, so strictly speaking it provides an upper bound on $C_k(P)$. The quantity $C_k(P)$ as defined in Eq. (22) is given by

$$C_k(P) = \inf_{\substack{T \in \mathcal{T}^{(\text{loc})} \\ Q \in \mathcal{E}_k}} D[P\|T(Q)], \quad (24)$$

so we need to optimize the choice of $Q \in \mathcal{E}_k$ and that of $T \in \mathcal{T}^{(\text{loc})}$ simultaneously. We here proceed iteratively. Given a test distribution $P(\cdot)$ we first find a Hamiltonian $H \in \mathcal{Q}_k$, parameterized by a set $\lambda^{(1)}$, so that $Q_{\lambda^{(1)}}$ is the best approximation of P in \mathcal{E}_k . As a next step we then find parameters $\omega^{(1)}$ which minimize the distance $D[P\|T_{\omega^{(1)}}(Q_{\lambda^{(1)}})]$, i.e. we find the point in the local orbit of $Q_{\lambda^{(1)}}$ closest to P . We then turn to an optimization of the k -particle Hamiltonian again, keeping $T_{\omega^{(1)}}$ fixed while optimizing $Q \in \mathcal{E}_k$. This is to say, find a set of parameters $\lambda^{(2)}$ such that $D[P\|T_{\omega^{(1)}}(Q_{\lambda^{(2)}})]$ is minimized (subject to the constraint $Q_{\lambda^{(2)}} \in \mathcal{E}_k$). Subsequently, we optimize the local transformation again, and find a new set of parameters $\omega^{(2)}$ minimizing $D[P\|T_{\omega^{(2)}}(Q_{\lambda^{(2)}})]$. This procedure is then iterated. Further details on the exact implementation can be found in the Appendix.

V. APPLICATION TO COUPLED CHAOTIC MAPS

In this section we apply the proposed measures of complexity to a specific case of an interacting multi-particle system. In particular we consider a coupled set of N discrete-time maps of the form [22]

$$\begin{aligned} x_i(t+1) &= (1-\varepsilon) f[x_i(t)] \\ &\quad + \frac{\varepsilon}{N-1} \sum_{j \neq i} f[x_j(t)], \end{aligned} \quad (25)$$

where $i, j = 1, \dots, N$. Similar arrangements were studied in Ref. [10], in particular different types of underlying adjacency networks were addressed. We here focus on the all-to-all coupling described in Eq. (25). The parameter $\varepsilon \in [0, 1]$ describes the strength of the interaction between the maps. For $\varepsilon = 0$ the N maps are fully uncoupled, for $\varepsilon > 0$ they become increasingly more coupled. The chaotic maps we have used are the tent map, given by

$$f(x) = \begin{cases} 2x & 0 \leq x \leq 1/2 \\ 2-2x & 1/2 \leq x \leq 1 \end{cases}, \quad (26)$$

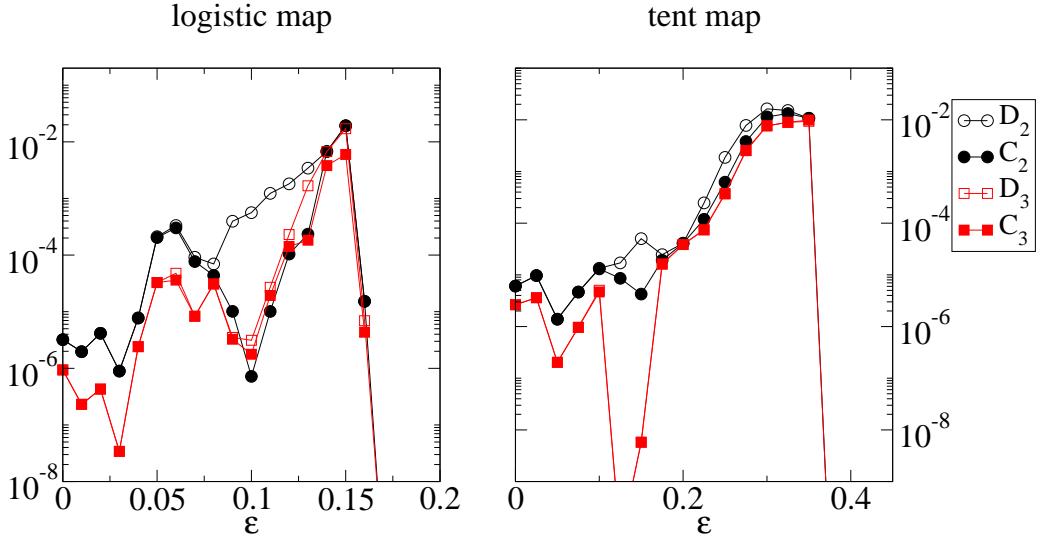


FIG. 3: (color online). Complexity measures C_k and D_k ($k = 2, 3$) for sets of $N = 4$ coupled logistic maps (left-hand panel), and $N = 4$ coupled tent maps (right-hand panel). Coarse-graining is performed at threshold $\Theta = 1/2$. Open symbols show the distances D_k from exponential families \mathcal{E}_k of probability measures generated by k -particle Hamiltonians, filled symbols are upper bounds on the distance from the local orbits of \mathcal{E}_k (see text for details).

and the logistic map defined by

$$f(x) = 4x(1-x). \quad (27)$$

For both maps a symbolic dynamics of binary symbols is derived from the continuous variables $x_i(t) \in [0, 1]$ via a simple coarse graining procedure of the type [23]

$$\sigma_i(t) = \begin{cases} 0 & x_i(t) \leq \Theta \\ 1 & x_i(t) > \Theta \end{cases}. \quad (28)$$

The parameter $\Theta \in (0, 1)$ is here a coarse-graining threshold.

To obtain the probability distribution, we start with a random configuration of the x_i and iterate the coupled map 10^6 times. We disregard the first 10^5 iterations and use the remaining iterations to determine our probability distribution.

A. Previous results

For completeness we will first re-iterate some of the results of Ref. [10]. Varying the value of the coupling strength ε we run simulations of a set of $N = 6$ coupled tent and logistic maps. For the choice $\Theta = 1/2$ we obtain the corresponding symbolic dynamics, and measure the resulting stationary distribution $P(\sigma)$, where $\sigma = (\sigma_1, \dots, \sigma_6)$. We then use the iterative projection algorithm proposed in Ref. [10] (and discussed in more detail in the Appendix), to obtain $D_k(P)$ for $k = 0, \dots, 6$. We then construct $I_k = D_{k-1} - D_k$ ($k = 1, \dots, 6$) from the outcome of the iterative projection. Results are shown

in Fig. 2 (red curves). As reported in Ref. [10] the system of coupled maps is driven toward a state of synchronized chaos if the coupling strength is sufficiently large ($\varepsilon \gtrsim 0.45$ for the system of tent maps, $\varepsilon \gtrsim 0.2$ for the logistic maps). In this regime the resulting distribution $P(\sigma)$ is found to be in \mathcal{E}_2 , and accordingly we have $I^{(k)} = 0$ for $k \geq 3$. With the exception of $I^{(2)}$, which is seen to be monotonically increasing in ε all other measures $I^{(k)}$ attain maxima at $\varepsilon \approx 0.35$ for the system of coupled tent maps, indicating that the most complex regime occurs just below the synchronization threshold [10]. A similar observation is made for the system of logistic maps, see the red curves in the lower six panels of Fig. 2.

B. Application of the modified complexity measure to coupled maps

Results for the modified complexity measures C_k are shown in Fig. 3. Realistic computing resources at present only allow us to study relatively small systems with $N = 4$ particles, as our implementation of the numerical computation of the improved measure C_k is significantly more demanding in computing time as that of the measures D_k . As the original complexity measure the modified measures are well able to detect the onset of synchronization. This is unsurprising as $D_k = 0$ implies $C_k = 0$.

For the cases we have studied here the functional dependence of C_k broadly follows that of D_k . For $k = 3$ in particular our numerical results for the two measures are essentially indistinguishable up to minor deviations. As seen in Fig. 3 the modified measure

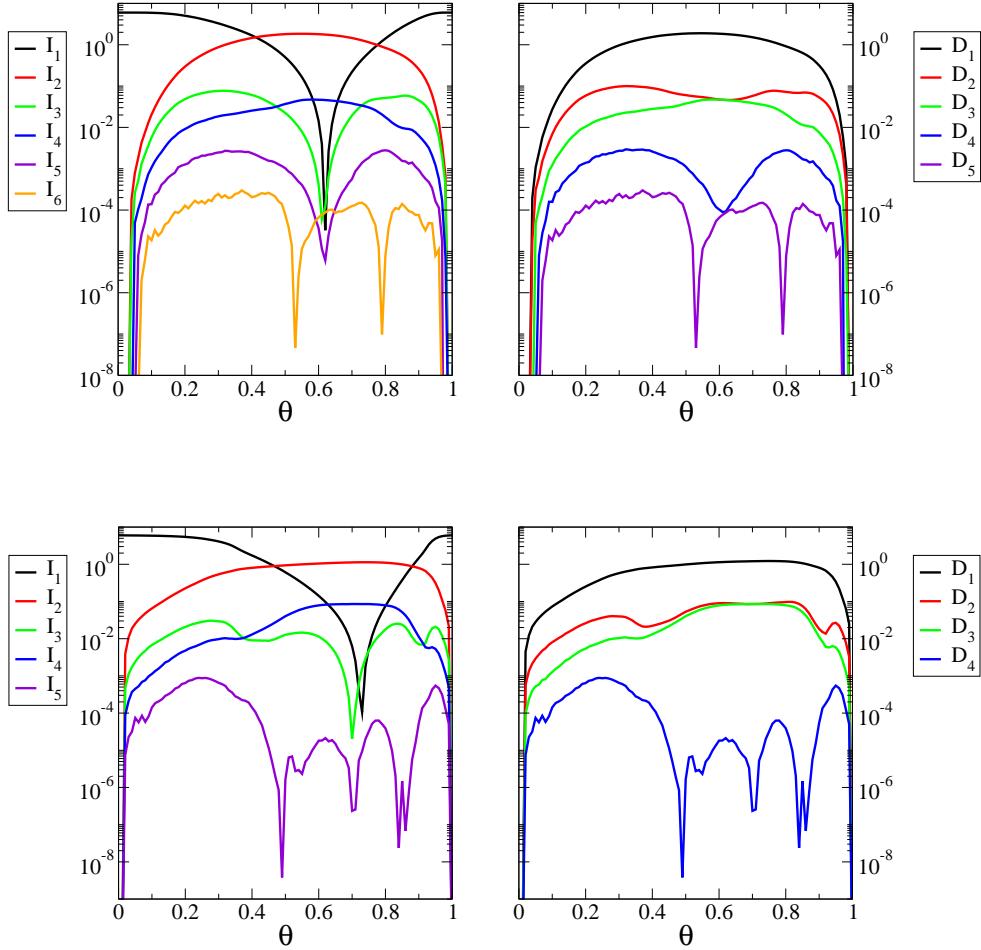


FIG. 4: (color online). Complexity measures I_k and distances D_k ($k = 1, \dots, 6$) in a system of $N = 6$ coupled tent maps (upper two panels) and for $N = 6$ logistic maps (lower two panels) as functions of the coarse-graining thresholds Θ . The coupling strength is fixed at $\varepsilon = 0.34$ for the tent map and at $\varepsilon = 0.17$ for the logistic map. All data is obtained single runs of the coupled maps, started at a random initial condition. The iterative projection algorithm is run for 1000 iterations.

of complexity can on occasion deviate substantially from the original one, see for example the data points near $\varepsilon = 0.1$ in the case of coupled logistic maps (left-hand panel of the figure). It is here important to keep in mind that the algorithm for the calculation of C_k is not guaranteed to converge. This, along with the relatively high costs in computing time, are clear drawbacks of the complexity measure we propose here. Future work may therefore address improved algorithms for computing these measures, and/or more detailed comparison with the complexity measures proposed by [10]. Should it turn out that both sets of measures give broadly the same results, then it may well be appropriate to continue to work with the D_k (or I_k), despite the lack of a non-increasing nature under local transformations.

C. Dependence on single threshold

As final point of our study we have investigated the influence of the quantitative value of the coarse graining threshold Θ on the complexity measures D_k and I_k . It is here important to be aware that the coarse graining procedure, mapping a real-valued variable $x_i(t)$ in the unit interval onto a discrete degree of freedom $\sigma_i(t) \in \{0, 1\}$, can be considered as a local operation as well, but it is a local procedure in the construction of the probability distribution, and not a local transformation applied after constructing the probability distribution.⁴ In performing the coarse graining each particle is treated independently from the rest of the system, and hence the applied opera-

⁴ Therefore, one cannot set the strict requirement for any complexity measure to be invariant under this transformation.

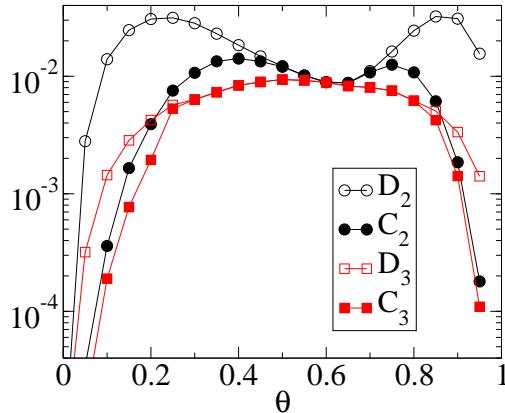


FIG. 5: (color online). Complexity measures C_k and D_k ($k = 2, 3$) for sets of $N = 4$ coupled tent maps at fixed coupling parameter $\varepsilon = 0.34$ and at varying coarse-graining threshold Θ . Open symbols show the distances D_k from exponential families \mathcal{E}_k of probability measures generated by k -particle Hamiltonians, filled symbols are upper bounds on the distance from the local orbits of \mathcal{E}_k .

tion is local.⁵

As seen in Fig. 2 the behaviour of the $\{I_k\}$ as functions of the coupling strength can depend considerably on the choice of the local threshold. All curves, except for those of I_2 for the case of coupled tent maps, show strong qualitative changes as Θ is varied. This is also confirmed in Fig. 4 where we show the Θ -dependence of I_k and D_k explicitly at a fixed coupling strength ε , chosen below, but close to the onset of synchronisation. Changing the numerical value of the threshold can result in changes of the I_k and D_k by orders of magnitude, and as seen in the left-hand panels of Fig. 4 it can also change the relative ordering of the I_k , therefore affecting their interpretation as measures of complexity.

In Fig. 5 we have tested the extent to which the use of the modified complexity measures C_k can remedy these effects. Again, for the three-particle correlation, we find that using local orbits instead of exponential families alone does not have any significant effect (we find that $C_3 \approx D_3$ for a broad range of values of the coarse-graining threshold Θ). The functional dependence of C_2 on the threshold parameter Θ does however appear to be weaker than that of the measure D_2 , we detect more variation of D_2 for intermediate values of Θ , where as the numerical estimate of C_2 varies only by comparably small amount. While this may provide some indication that the modified measures exhibit a lesser dependence on the details of the

coarse-graining procedure that the original measure of complexity, more work is required to test whether this continues to hold for other systems and/or systems with a larger number of particles

VI. CONCLUSION

In summary, we have investigated measures of complexity and emergence for classical interacting particle systems. Based on concepts from information geometry such measures have recently been proposed [10]. The main idea behind these measures is to consider the distance of a given probability distribution from the set of distributions generated by k -particle Hamiltonians, that is to say the set of distributions which can be factorized into a product of k -partite distributions. We have shown that such measures are generally not invariant under local transformations, acting on individual particles. In particular we have demonstrated that, somewhat counterintuitively, local transformations can increase the degree of multi-particle complexity in this measure, generating for example 3-particle correlations from 2-particle correlations. Similarly integrating out an individual degree of freedom can increase the complexity of the remaining marginal. Recent work by Zhou [16] has revealed analogous findings in the quantum realm, our work here demonstrates that these effects are not intrinsically quantum, for example related to the non-commutative nature of quantum mechanics. Instead they are also seen in classical systems.

In order to remedy these undesired properties of existing complexity measures, we have devised a modification, and propose to consider the distance from local orbits of exponential families instead of the distance from the families themselves. These orbits are manifestly invariant under local transformations of the type we have defined, and as a consequence the resulting complexity measure can only be reduced by applying a local transformation, but not increased. We have devised a numerical scheme with which to calculate upper bounds for this complexity measure, results are presented for dynamical systems composed of multiple interacting chaotic maps.

Finally we have investigated how the choice of local coarse-graining thresholds affects the complexity measures proposed by Ref. [10]. The coarse graining procedure here constitutes a local manipulation as well, albeit different from the type we consider above. Numerical results indicate that the choice of coarse graining threshold can have significant effects on the resulting estimates of complexity, hence care needs to be taken in turning the underlying continuous degrees of freedom into symbolic language on the coarse-grained level.

While the modified complexity measure we put forward here appears to exhibit more convenient be-

⁵ The following procedure would constitute an example of a nonlocal coarse-graining: set $\sigma_i = 0$ if $x_i(t)x_{i+1}(t) \leq 1/2$ and $\sigma_i = 1$ otherwise. The expression $i+1$ is to be taken as ‘mod n ’.

haviour under local transformation that the ones proposed in Ref. [10] it is important for keep in mind that the algorithm we are able to propose produces upper bounds for the required complexity measures, but at present we are unable to decide whether or not it converges to the desired quantities asymptotically. More work along these lines is required. Additionally the numerical scheme we use here is limited to small systems, hence there is significant room for improvement. This is however beyond the scope of the current work which aims mainly to identify the key properties of the complexity measures we have discussed, to relate them to measures based on exponential families and to discuss their behaviour under local transformations. Our work is therefore mostly of a conceptual nature, and we hope our results will stimulate further work towards efficient algorithms for complexity measures with the local invariance property. Additional future lines of research may also address geometric concepts of distances other than those based on the Kullback-Leibler divergence. Finally, it would also be interesting to study the quantum analog of exponential families and their local orbits in more detail. Work along these lines is in progress.

Acknowledgments

We thank Nihat Ay, Jürgen Jost, Robert MacKay, Sönke Niekamp, and Eckehard Olbrich for discussions. This work is supported by the Royal Society (reference JP090467). TG would like to thank Research Councils UK for support (RCUK reference EP/E500048/1). OG acknowledges support by the FWF (START prize Y376-N16) and the EU (Marie Curie CIG 293993/ENFOQI).

VII. APPENDIX: ALGORITHMS

This Appendix contains details on the two main algorithms used in this work. In Sec. VII A we provide the precise steps with which to compute $D_k(P)$ for a given distribution P , i.e. the distance from the set of distributions generated by k -particle Hamiltonians. In Sec. VII B we explain in more detail how to compute an upper bound for $C_k(P)$, the distance of P from the local orbit of \mathcal{E}_k .

A. Iterative projection algorithm

Our presentation here closely follows that of Ref. [14], see also Ref. [10]. Given an (empirical) probability distribution $P(\cdot)$ on $\Omega = \{0, 1\}^N$, and an integer $k \leq N$ (the order of interaction at which we

are approximating), the algorithm to find $D_k(P)$ is as follows:

1. *Computation of marginals:* For each subset $A \subset V$ with k elements ($|A| = k$), compute the following

$$\alpha_A(\boldsymbol{\sigma}) = \sum_{\boldsymbol{\sigma}' : \pi_A(\boldsymbol{\sigma}') = \pi_A(\boldsymbol{\sigma})} P(\boldsymbol{\sigma}'). \quad (29)$$

Here $\pi_A(\boldsymbol{\sigma})$ denotes the projection of $\boldsymbol{\sigma}$ onto A , i.e. $\pi_A(\boldsymbol{\sigma}) = (\sigma_i)_{\{i \in A\}}$. The quantity $\alpha_A(\boldsymbol{\sigma})$ only depends on the components σ_i of $\boldsymbol{\sigma}$ with index $i \in A$, i.e. on the k -variables $(\sigma_i)_{i \in A}$.

2. *Initialisation:* Initialise Q as the flat distribution over Ω : $Q(\boldsymbol{\sigma}) = 1/(2^N)$ for all $\boldsymbol{\sigma}$.
3. *Improve current approximation:* Run through all sets $A \subset V$ with $|A| = k$ and update $Q(\cdot)$ as follows:

$$Q^{\text{new}}(\boldsymbol{\sigma}) = c_A(\boldsymbol{\sigma})Q(\boldsymbol{\sigma}), \quad (30)$$

where

$$c_A(\boldsymbol{\sigma}) = \frac{\alpha_A(\boldsymbol{\sigma})}{\sum_{\boldsymbol{\sigma}' : \pi_A(\boldsymbol{\sigma}') = \pi_A(\boldsymbol{\sigma})} Q(\boldsymbol{\sigma}')}. \quad (31)$$

4. *Update and iterate:* Replace $Q(\cdot) \rightarrow Q^{\text{new}}(\cdot)$ and goto 3.

B. Approximate scheme to calculate distance from a local orbit

Given a test distribution $P(\cdot)$ the algorithm proceeds as follows:

1. Find a Hamiltonian $H \in \mathcal{Q}_k$, parameterized by a set $\boldsymbol{\lambda}^{(1)}$, so that $Q_{\boldsymbol{\lambda}^{(1)}} = Z^{-1} \exp(H)$ is the best approximation of P in \mathcal{E}_k . This can for example be done using the iterative projection algorithm described above.
2. *Optimization of the local transformation for a given Hamiltonian:* Given a Hamiltonian parameterized by $\boldsymbol{\lambda}^{(m)}$ find parameters $\boldsymbol{\omega}^{(m)}$ which minimize the distance $D[P \| T_{\boldsymbol{\omega}^{(m)}}(Q_{\boldsymbol{\lambda}^{(m)}})]$, i.e. we find the point in the local orbit of $Q_{\boldsymbol{\lambda}^{(m)}}$ closest to P . The optimization of the parameter set $\boldsymbol{\omega} = (a_1, b_1, \dots, a_N, b_N)$ is carried out in an iterative manner, i.e. we first optimize a_1 , then b_1 , then a_2 , and so on, keeping previously optimized parameters fixed. This procedure is iterated a number of times (typically in excess of 10 sweeps). Each parameter optimization is carried out using an iterative Monte Carlo procedure, based on first randomly choosing test

values in the interval $[0, 1]$ and then choosing a sequence of nested intervals with decreasing the range from which more finely spaced subsequent test values are drawn.

3. *Optimization of the Hamiltonian given a local transformation:* Given a local transformation defined by parameters $\boldsymbol{\omega}^{(m)}$ find a set of parameters $\boldsymbol{\lambda}^{(m+1)}$ such that $D[P||T_{\boldsymbol{\omega}^{(m)}}(Q_{\boldsymbol{\lambda}^{(m+1)}})]$ is minimized (subject to the constraint $Q_{\boldsymbol{\lambda}^{(m+1)}} \in \mathcal{E}_k$). Similar to the procedure outlined above for the optimization of $\boldsymbol{\omega}$ we optimize one of the parameters $\boldsymbol{\lambda}$ at a time, carrying out a set of typically 10 – 20 or more sweeps over all parameters. Any one parameter is optimized by applying a deterministic search algorithm on the interval $[-10, 10]$ (with a suitable discretization). The range of each parameter is therefore effectively truncated. The constraint $Q_{\boldsymbol{\lambda}^{(m+1)}} \in \mathcal{E}_k$ is taken into account by pre-setting the coefficients of ℓ -particle terms, $\ell > k$ to zero. For a three-particle system for example we have $\boldsymbol{\lambda} = (\alpha_1, \alpha_2, \alpha_3, \beta_{12}, \beta_{23}, \beta_{13}, \gamma)$, see

Eq. (15). If the wish to compute differences say from \mathcal{E}_2 we would set $\gamma = 0$ from the start, and only optimize the remaining entries in $\boldsymbol{\lambda}$.

4. This procedure is then iterated, i.e., goto 2.

We would like to stress that no claim is made that this algorithm provides an exact result for C_k , this may not even be the case in the limit of an infinite number of iterations. The numerical scheme is an approximate procedure, providing at least an *upper bound* on C_k . It is also interesting to note that the algorithm contains a certain stochastic element (rooted in the Monte-Carlo optimization of the parameters $\boldsymbol{\omega}$ as described above). We have found that it can be beneficial to allow occasional increases in the estimates of distance (i.e. ‘uphill’ motion), as this prevents dynamical arrest in local minima. Results reported are the minimal estimate of distance obtained during any one run of the algorithm, and not necessarily the distance estimate at the end of the nested set of iterations.

[1] J. P. Crutchfield, *Physica D* **75**, 11 (1994).
[2] J. P. Crutchfield and K. Young, *Phys. Rev. Lett.* **63**, 105 (1989).
[3] Y. Bar-Yam, *Complexity* **9**, 15 (2004).
[4] D.J. Chalmers, *Strong and Weak Emergence, The Re-Emergence of Emergence*, Oxford University Press, Oxford UK (2002).
[5] R. C. Ball, M. Diakonova, R. S. MacKay, *Adv. Compl. Sys.* **13**, 327 (2010).
[6] M. Diakonova and R. S. MacKay, *Int. J. Bif. and Chaos* **21** 2297 (2011)
[7] R.S. MacKay, Indecomposable coupled map lattices with non-unique phase, in: Dynamics of coupled map lattices and of related spatially extended systems, eds Chazottes J-R, Fernandez B, *Lect Notes Phys* **671**, 65-94 (Springer, 2005).
[8] R.S. MacKay, *Nonlinearity* **21**, T273 (2008).
[9] K. Lindgren and M. G. Nordahl, *Complex Systems* **2** 409 (1988).
[10] T. Kahle, E. Olbrich, J. Jost, and N. Ay, *Phys. Rev. E* **79**, 026201 (2009).
[11] N. Ay, E. Olbrich, N. Bertschinger, and J. Jost. A Unifying Framework for Complexity Measures of Finite Systems, *Proceedings ECCS’06* (2006). Santa Fe Institute Working Paper 06-08-028.
[12] S. Amari, *IEEE Trans. Inf. Theory* **47**, 1701 (2001).
[13] E. Olbrich, T. Kahle, N. Bertschinger, N. Ay, and J. Jost, arXiv:0912.4450.
[14] L. Steiner and T. Kahle *Computing Information Projections Iteratively with CIPI*, available at <https://github.com/tom111/cipi>. (accessed August 2010)
[15] For reviews see M. Plenio and S. Virmani, *Quant. Inf. Comp.* **7**, 1 (2007). R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, *Rev. Mod. Phys.* **81**, 865 (2009); O. Gühne and G. Tóth, *Phys. Reports* **474**, 1 (2009).
[16] D. L. Zhou, *Phys. Rev. A* **80**, 022113 (2009).
[17] N. Linden, S. Popescu, and W. K. Wootters, *Phys. Rev. Lett.* **89**, 207901 (2002).
[18] D. L. Zhou, *Phys. Rev. Lett.* **101**, 180505 (2008).
[19] D. L. Zhou, B. Zeng, Z. Xu, and L. You, *Phys. Rev. A* **74**, 052110 (2006).
[20] T.M. Cover and J.A. Thomas *Elements of Information Theory* John Wiley & Sons (2006).
[21] N. Ay and A. Knauf, *Kybernetika* **42**, 517 (2007).
[22] K. Kaneko, *Chaos* **2**, 279 (1992).
[23] S. Jalan, F. Atay, and J. Jost, Detecting global properties of coupled dynamics using local symbolic dynamics. *Chaos* **16**, 1 (2006).